



5-32683P7
The
Patent
Office



INVESTOR IN PEOPLE

The Patent Office
Concept House
Cardiff Road
Newport
South Wales
NP10 8QQ



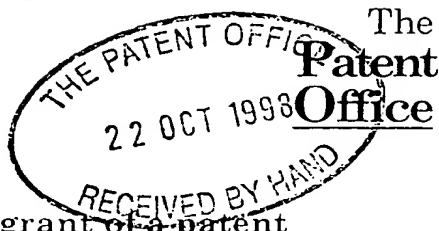
I, the undersigned, being an officer duly authorised in accordance with Section 74(1) and (4) of the Deregulation & Contracting Out Act 1994, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

In accordance with the rules, the words "public limited company" may be replaced by p.l.c., plc, P.L.C. or PLC.

Re-registration under the Companies Act does not constitute a new legal entity but merely subjects the company to certain additional company law rules.

Signed *Andrew Gersey*
Dated 2 September 1999



1/77

Request for grant of a patent

(See the notes on the back of this form. You can also get an explanatory leaflet from the Patent Office to help you fill in this form)

22 OCT 1998

The Patent Office
Cardiff Road
Newport
Gwent NP9 1RH

1. Your reference **S-30683/P1**

2. Patent application number
(The Patent Office will fill in this part)

9823098.0

3. Full name, address and postcode of the or of each applicant
(underline all surnames)

**NOVARTIS AG
SCHWARZWALDALLEE 215
4058 BASEL
SWITZERLAND**

Patent ADP number (if you know it)

If the applicant is a corporate body, give the country/state of its incorporation

SWITZERLAND

712 5487002

4. Title of invention

Organic compounds

5. Name of your agent (If you have one)

"Address for service" in the United Kingdom to which all correspondence should be sent
(including the postcode)

**B.A. YORKE & CO.
CHARTERED PATENT AGENTS
COOMB HOUSE, 7 ST. JOHN'S ROAD
ISLEWORTH
MIDDLESEX TW7 6NH**

Patents ADP number (if you know it)

1800001

6. If you are declaring priority from one or more earlier patent applications, give the country and the date of filing of the or of each of these earlier applications and (if you know it) the or each application number

Country

Priority application number
(if you know it)

Date of filing
(day/month/year)

7. If this application is divided or otherwise derived from an earlier UK application, give the number and the filing date of the earlier application

Number of earlier
application

Date of filing
(day/month/year)

8. Is a statement of inventorship and of right to grant of a patent required in support of this request? (Answer 'Yes' if:

Yes

a) any applicant named in part 3 is not an inventor, or

b) there is an inventor who is not named as an applicant, or

c) any named applicant is a corporate body.

(see note (d))

Patents Form 1/77

9. Enter the number of sheets for any of the following items you are filing with this form. Do not count copies of the same document

Continuation sheets of this form

Description 41

Claim(s) 2

Abstract -

Drawing(s) -

10. If you are also filing any of the following, state how many against each item.

Priority documents


Translations of priority documents

Statement of inventorship and right to grant of a patent (*Patents Form 7/77*)

Request for preliminary examination and search (*Patents Form 9/77*)

Request for substantive examination (*Patents Form 10/77*)

Any other documents
(please specify)

11. I/We request the grant of a patent on the basis of this application
- | Signature | Date |
|---------------------------------------------------------------------------------------------------------|----------|
| 
B.A. Yorke & Co. | 22.10.98 |
12. Name and daytime telephone number of person to contact in the United Kingdom
- | |
|------------------|
| Mrs. E. Cheetham |
| 0181 560 5847 |

Warning

After an application for a patent has been filed, the Comptroller of the Patent Office will consider whether publication or communication of the invention should be prohibited or restricted under Section 22 of the Patents Act 1977. You will be informed if it is necessary to prohibit or restrict your invention in this way. Furthermore, if you live in the United Kingdom, Section 23 of the Patents Act 1977 stops you from applying for a patent abroad without first getting written permission from the Patent Office unless an application has been filed at least 6 weeks beforehand in the United Kingdom for a patent for the same invention and either no direction prohibiting publication or communication has been given, or any such direction has been revoked.

Notes

- If you need help to fill in this form or you have any questions, please contact the Patent Office on 0645 500505.
- Write your answers in capital letters using black ink or you may type them.
- If there is not enough space for all the relevant details on any part of this form, please continue on a separate sheet of paper and write "see continuation sheet" in the relevant part(s). Any continuation sheet should be attached to this form.
- Once you have filled in the form you must remember to sign and date it.
- For details of the fee and ways to pay please contact the Patent Office.

Organic Compounds

The present invention relates to vegetative reproduction of plants and plant cells. In particular the invention relates to a method for increasing the probability of vegetative reproduction *in vivo* through seeds or *in vitro* by somatic embryogenesis. Apomictic seeds resulting therefrom, and the plants and progeny obtained through germination of such seeds are further subject matters of the invention.

Vegetative, non-sexual reproduction through seeds also called apomixis, is a genetically controlled reproductive mechanism of plants found in some polyploid non-cultivated species. Two types of apomixis, gametophytic or non-gametophytic, can be distinguished. In gametophytic apomixis - of which there are two types, namely apospory and diplospory - multiple embryo sacs typically lacking antipodal nuclei are formed, or else megasporogenesis in the embryo sac takes place. In non-gametophytic apomixis also called adventitious embryony, a somatic embryo develops directly from the cells of the embryo sac, ovary wall or integuments. Somatic embryos from surrounding cells invade the sexual ovary, one of the somatic embryos out-competes the other somatic embryos and the sexual embryo, and utilizes the produced endosperm.

Engineering apomixis to a controllable, more reproducible trait would provide many advantages in plant improvement and cultivar development in case that sexual plants are available as crosses with the apomictic plant. The Somatic Embryogenesis Receptor Kinase (SERK) is known to be involved in the formation of extraneous embryos from sporophytic cells which can result in apomictic seeds.

Apomixis would provide for true-breeding, seed propagated hybrids. Moreover, apomixis could shorten and simplify the breeding process so that selfing and progeny testing to produce and/or stabilize a desirable gene combination could be eliminated. Apomixis would provide for the use as cultivars of genotypes with unique gene combinations since apomictic genotypes breed true irrespective of heterozygosity. Genes or groups of genes could thus be "pyramided and "fixed" in super genotypes. Every superior apomictic genotype from a sexual-apomictic cross would have the potential to be a cultivar. Apomixis would allow plant breeders to develop cultivars with specific stable traits for such characters as height, seed and forage quality and maturity.

Breeders would not be limited in their commercial production of hybrids by (i) a cytoplasmic-nuclear interaction to produce male sterile female parents or (ii) the fertility restoring capacity of a pollinator. Almost all cross-compatible germplasm could be a potential parent to produce apomictic hybrids.

Apomixis would also simplify commercial hybrid seed production. In particular, (i) the need for physical isolation of commercial hybrid production fields would be eliminated; (ii) all available land could be used to increase hybrid seed instead of dividing space between pollinators and male sterile lines; and (iii) the need to maintain parental line seed stocks would be eliminated.

The potential benefits to accrue from the production of seed *via* apomixis are presently unrealized, to a large extent because of the problem of engineering apomictic capacity into plants of interest. The present invention teaching introduction of proteins acting in the signal transduction cascade triggered by SERK provides a further step to the solution of that problem in that it improves vegetative reproduction *in vivo* through seeds and *in vitro* by somatic embryogenesis.

In the following the term "gene" refers to a coding sequence and associated regulatory sequences. The coding sequence is transcribed into RNA, which depending on the specific gene, will be mRNA, rRNA, tRNA, snRNA, sense RNA or antisense RNA. Examples of regulatory sequences are promoter sequences, 5' and 3' untranslated sequences and termination sequences. Further elements that may be present are, for example, introns.

A "promoter" is a DNA sequence initiating transcription of an associated DNA sequence. Depending on the specific promoter region it may also include elements that act as regulators of gene expression such as activators, enhancers, and/or repressors.

A regulatory DNA sequence such as promoter is said to be "operably linked to" or "associated with" a DNA sequence that codes for an RNA or a protein, if the two sequences are situated such that the regulatory DNA sequence affects expression of the coding DNA sequence.

The term "expression" refers to the transcription and/or translation of an endogenous gene or a transgene in plants.

Expression "in the vicinity of the embryo sac" is considered to mean expression in carpel, integuments, ovule, ovule premordium, ovary wall, chalaza, nucellus, funicle or placenta. The skilled man will recognize that the term "integuments" can include tissues which are derived therefrom, such as endothelium. "Embryogenic" defines the capability of cells to develop into an embryo under permissive conditions. It will be appreciated that the term "in an active form" includes proteins which are truncated or otherwise mutated with the proviso that they still increase the probability of vegetative reproduction whether or not in doing this they interact with the signal transduction components that they otherwise would in the tissues in which they are normally present.

"Marker genes" encode a selectable or screenable trait. Thus, expression of a "selectable marker gene" gives the cell a selective advantage which may be due to their ability to grow in the presence of a negative selective agent, such as an antibiotic or a herbicide, compared to the growth of non-transformed cells. The selective advantage possessed by the transformed cells, compared to non-transformed cells, may also be due to their enhanced or novel capacity to utilize an added compound as a nutrient, growth factor or energy source. Selectable marker gene also refers to a gene or a combination of genes whose expression in a plant cell gives the cell both, a negative and a positive selective advantage. On the other hand a "screenable marker gene" does not confer a selective advantage to a transformed cell, but its expression makes the transformed cell phenotypically distinct from untransformed cells.

The term "plant" refers to any plant, but particularly seed plants.

The term "plant cell" describes the structural and physiological unit of the plant, and comprises a protoplast and a cell wall. The plant cell may be in form of an isolated single cell (such as stomatal guard cells) or a cultured cell, or as a part of higher organized unit such as, for example, a plant tissue, or a plant organ.

The term "plant material" includes leaves, stems, roots, emerged radicles, flowers or flower parts, petals, fruits, pollen, pollen tubes, anther filaments, ovules, embryo sacs, egg cells, ovaries, zygotes, embryos, zygotic embryos *per se*, somatic embryos, hypocotyl sections,

apical meristems, vascular bundles, pericycles, seeds, cuttings, cell or tissue cultures, or any other part or product of a plant

The following solutions are provided by the present invention:

- A method for increasing the probability of vegetative reproduction of a new plant generation comprising transgenically expressing a gene encoding a protein acting in the signal transduction cascade triggered by the Somatic Embryogenesis Receptor Kinase (SERK);
- said method wherein the encoded protein physically interacts with SERK;
- said method wherein the protein is a member of the family of Squamosa-promoter Binding Protein (SBP) transcription factors or 14-3-3 type lambda proteins;
- said method wherein the protein has the amino acid sequence given in SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, SEQ ID NO: 10, SEQ ID NO: 12, SEQ ID NO: 14, or SEQ ID NO: 16, or an amino acid sequence having a component sequence of at least 150 amino acids length which after alignment reveals at least 40% identity with SEQ ID NO: 12 or SEQ ID NO: 16;
- said method increasing the probability of vegetative reproduction through seeds (apomixis);
- said method wherein the seeds result from non-gametophytic apomixis;
- said method wherein the encoded protein is transgenically expressed in the vicinity of the embryo sac;
- said method increasing the probability of *in vitro* somatic embryogenesis;
- said method wherein expression of the gene is under control of the SERK gene promoter, the carrot chitinase DcEP3-1 gene promoter, the *Arabidopsis* AtChitIV gene promoter, The *Arabidopsis* LTP-1 gene promoter, The *Arabidopsis* bel-1 gene promoter, the petunia fbp-7 gene promoter, the *Arabidopsis* ANT gene promoter or the promoter of the O126 gene of *Phalaenopsis*;
- a gene encoding a protein having the amino acid sequence given in SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, SEQ ID NO: 10, SEQ ID NO: 12, SEQ ID NO: 14, or SEQ ID NO: 16, or an amino acid sequence having a component sequence of at least 150 amino acids length which after alignment reveals at least 40% sequence identity with SEQ ID NO: 12 or SEQ ID NO: 16;

- said gene having the nucleotide sequence given in SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 9, SEQ ID NO: 11, SEQ ID NO: 13, or SEQ ID NO: 15;
- said gene wherein the nucleotide sequence is modified in that known mRNA instability motifs or polyadenylation signals are removed and/or codons which are preferred by the plant into which the DNA is to be inserted are used;
- a plant or plant cell transgenically expressing said gene; and
- a plant or plant cell obtainable by the method according to the present invention.

According to the present invention there is provided a method for increasing the probability of vegetative reproduction of a new plant generation, for example by producing apomictic seeds or generating somatic embryos under *in vitro* conditions, comprising transgenically expressing a gene encoding a protein acting in the signal transduction cascade triggered by the Somatic Embryogenesis Receptor Kinase (SERK). This is achieved by

- (i) transforming plant material with a nucleotide sequence encoding said protein,
- (ii) regenerating transformed plant material into plants, or carpel-containing parts thereof, and
- (iii) expressing the sequence in the vicinity of the embryo sac.

A further embodiment of the invention relates to genes encoding proteins acting in the signal transduction cascade triggered by the Somatic Embryogenesis Receptor Kinase (SERK) the presence of which in an active form in a cell, or membrane thereof, renders said cell embryogenic.

The gene to be expressed preferably encodes a protein physically interacting with SERK. Specific examples of SERK-interacting proteins are members of the family of Squamosa-promoter Binding Protein (SBP) transcription factors (Klein et al, Mol Gen Genet 250: 7-16, 1996). These proteins are able to interact specifically with DNA through a conserved domain of 70 to 90, preferably 79 amino acid residues, the SBP-box. Alignment of different SBP-box sequences generally reveals at least 50% and preferably more than 60% or more than 70 % sequence identity. Within the SBP-box a remarkable arrangement of cysteine and histidine residues can be recognized, which is reminiscent of zinc-fingers and probably involved in the recognition of specific promoter elements. A bipartite nuclear localization signal is placed at the C-terminal end of the SBP-box (Dingwall et al, Trends Biochem Sci 16: 478-481, 1991). Both the N-terminal and the C-terminal domains of the SERK-

interacting SBP proteins are highly variable and are probably involved in regulation of protein activity. One of the possible SBP proteins is identical with SPL3 (SEQ ID NO: 5 and SEQ ID NO: 6), a gene involved in the floral transition and expressed in developing flower buds (Cardon et al, Plant Journal 12: 367-377, 1997).

Another class of SERK-interacting proteins are isoforms of the family of 14-3-3 proteins such as the 14-3-3 type lambda protein (Wu et al, Plant Physiol 114: 1421-1431, 1997; SEQ ID NO: 9 and SEQ ID NO: 10). A total of 10 different 14-3-3 proteins are present in *Arabidopsis* the different members being involved in intracellular signal transduction. They mediate signal transduction by binding to phosphoserine-containing proteins on specific binding motifs represented by conserved amino acid sequences like RxxS(p)xP (Yaffe et al, Cell 91: 961-971, 1997). A putative 14-3-3 interaction domain having the amino acid sequence RPPSQP is also found at position 391-396 of the *Arabidopsis* SERK protein, and at the corresponding aligned region of the *Daucus carota* SERK protein having the amino acid sequence RQPSEP providing SERK with a mechanism for a 14-3-3 mediated signal transduction.

A further class of SERK-interacting proteins is exemplified by SEQ ID NO: 11 (and SEQ ID NO: 12) and the NDR1 protein already described in the literature (Century et al, Science 278: 1963-1965, 1997). NDR1 is likely to encode a membrane-associated component in the signal transduction pathway downstream of pathogen-recognizing proteins. It was suggested that NDR1 might be a protein that interacts with many different receptors. SEQ ID NO: 6 represents a new member in this small family of proteins supposed to function in intracellular signal transduction mediated by transmembrane receptors.

SEQ ID NO: 13 encodes a SERK-interacting protein (SEQ ID NO: 14) with homology to a domain of *E.coli* aminopeptidase N and is expected to encode an *Arabidopsis* protease interacting with or activated by SERK.

The predicted amino acid sequence of the SERK-interacting protein of SEQ ID NO: 15 (SEQ ID NO: 16) has no homology with known gene products although there is a small not yet described family of related gene products in *Arabidopsis*.

Insofar as the the SERK-interacting proteins mentioned above and their corresponding genes are novel they constitute a further subject matter of the present invention.

Of course, genes similar to the ones described above can also be used. A similar gene is a gene having a nucleotide sequence complementary to the test sequence and capable of hybridizing to the inventive sequence. When the test and inventive sequences are double stranded the

nucleic acid constituting the test sequence preferably has a T_M within 20°C of that of the inventive sequence. In the case that the test and inventive sequences are mixed together and denatured simultaneously, the T_M values of the sequences are preferably within 10°C of each other. More preferably the hybridization is performed under stringent conditions, with either the test or inventive DNA preferably being supported. Thus either a denatured test or inventive sequence is preferably first bound to a support and hybridization is effected for a specified period of time at a temperature of between 50° and 70°C in double strength citrate buffered saline (SSC) containing 0.1% SDS followed by rinsing of the support at the same temperature but with a buffer having a reduced SSC concentration. Depending upon the degree of stringency required, and thus the degree of similarity of the sequences, at a particular temperature, - such as 60°C , for example - such reduced concentration buffers are typically single strength SSC containing 0.1% SDS, half strength SSC containing 0.1% SDS and one tenth strength SSC containing 0.1% SDS. Sequences having the greatest degree of similarity are those the hybridization of which is least affected by washing in buffers of reduced concentration. It is most preferred that the test and inventive sequences are so similar that the hybridization between them is substantially unaffected by washing or incubation in one tenth strength sodium citrate buffer containing 0.1% SDS.

The gene to be expressed may be modified in that known mRNA instability motifs or polyadenylation signals are removed or codons which are preferred by the plant into which the sequence is to be inserted may be used so that expression of the thus modified sequence in the said plant may yield substantially similar protein to that obtained by expression of the unmodified sequence in the organism in which the protein is endogenous.

The sequence variability of proteins with similar function suggests, that a number of amino acids can be replaced, inserted or deleted without altering a protein's function. The relationship between proteins is reflected by the degree of sequence identity between aligned amino acid sequences of individual proteins or aligned component sequences thereof.

Dynamic programming algorithms yield different kinds of alignments. In general there exist two approaches towards sequence alignment. Algorithms as proposed by Needleman and Wunsch and by Sellers align the entire length of two sequences providing a global alignment of the sequences. The Smith-Waterman algorithm on the other hand yields local alignments. A local alignment aligns the pair of regions within the sequences that are most

similar given the choice of scoring matrix and gap penalties. This allows a database search to focus on the most highly conserved regions of the sequences. It also allows similar domains within sequences to be identified. To speed up alignments using the Smith-Waterman algorithm both BLAST (Basic Local Alignment Search Tool) and FASTA place additional restrictions on the alignments.

Within the context of the present invention alignments are conveniently performed using BLAST, a set of similarity search programs designed to explore all of the available sequence databases regardless of whether the query is protein or DNA. Version BLAST 2.0 (Gapped BLAST) of this search tool has been made publicly available on the internet (currently <http://www.ncbi.nlm.nih.gov/BLAST/>). It uses a heuristic algorithm which seeks local as opposed to global alignments and is therefore able to detect relationships among sequences which share only isolated regions. The scores assigned in a BLAST search have a well-defined statistical interpretation. Particularly useful within the scope of the present invention are the blastp program allowing for the introduction of gaps in the local sequence alignments and the PSI-BLAST program, both programs comparing an amino acid query sequence against a protein sequence database, as well as a blastp variant program allowing local alignment of two sequences only. Said programs are preferably run with optional parameters set to the default values.

Sequence alignments using BLAST can also take into account whether the substitution of one amino acid for another is likely to conserve the physical and chemical properties necessary to maintain the structure and function of a protein or is more likely to disrupt essential structural and functional features. For example non-conservative replacements may occur at a low frequency and conservative replacements may be made between amino acids within the following groups:

- (i) Serine and Threonine;
- (ii) Glutamic acid and Aspartic acid;
- (iii) Arginine and Lysine;
- (iv) Asparagine and Glutamine;
- (v) Isoleucine, Leucine, Valine and Methionine;
- (vi) Phenylalanine, Tyrosine and Tryptophan
- (vii) Alanine and Glycine.

Such sequence similarity is quantified in terms of a percentage of positive amino acids, as compared to the percentage of identical amino acids and can help assigning a protein to the correct protein family in border-line cases.

Specific embodiments of the invention express a gene comprising a DNA sequence encoding a protein acting in the signal transduction cascade triggered by the Somatic Embryogenesis Receptor Kinase (SERK) and having the amino acid sequence depicted in SEQ ID NO: 2, 4, 6 or 8, or a protein similar thereto. By similar is meant a protein having a component sequence of at least 150 amino acids length which after alignment reveals at least 40% and preferably 50% or more sequence identity with another protein.

In order to obtain expression of the sequence in a regenerated plant and in particular the carpel thereof in a tissue specific manner the sequence is under expression control of an inducible or developmentally regulated promoter. It is preferred that the gene is expressed in the somatic cells of the embryo sac, ovary wall, nucellus, or integuments. As the endosperm within the apomictic seed results from fusion of polar nuclei within the embryo sac with a pollen-derived male gamete nucleus it is preferred that the sequence encoding the protein is expressed prior to fusion of the polar nuclei with the male gamete nucleus.

Typically promoters are a promoter which regulates expression of SERK genes *in planta*, the *Arabidopsis* ANT gene promoter, the promoter of the O126 gene from *Phalaenopsis*, the carrot chitinase DcEP3-1 gene promoter, the *Arabidopsis* AtChitIV gene promoter, the *Arabidopsis* LTP-1 gene promoter, the *Arabidopsis* bel-1 gene promoter, the petunia fbp-7 and fbp-11 gene promoters, the *Arabidopsis* AtDMC1 promoter, the pTA7001 inducible promoter. The DcEP3-1 gene is expressed transiently during inner integument degradation and later in cells that line the inner part of the developing endosperm. The AtChitIV gene is transiently expressed in the micropylar endosperm up to cellularisation. The LTP-1 promoter is active in the epidermis of the developing nucellus, both integuments, seed coat and early embryo. The bel-1 gene is expressed in the developing inner integument and the fbp-7 promoter is active during embryo sac development. The *Arabidopsis* ANT gene is expressed during integument development, and the O126 gene from *Phalaenopsis* is expressed in the mature ovule.

The promoters of the DcEP3-1 and the AtChit IV genes may be cloned and characterized by standard procedures. The gene encoding a protein of the SERK signal cascade is cloned behind the DcEP3-1, the AtChit IV or the AtLTP-1 promoters and transformed into *Arabidopsis*.

The ligation is performed in such a way that the promoter is operably linked to the sequence to be transcribed. This construct, which also contains known marker genes providing for selection of transformed material, is inserted into the T-DNA region of a binary vector such as pBIN19 and transformed into *Arabidopsis*. *Agrobacterium*-mediated transformation into *Arabidopsis* is performed by the vacuum infiltration or root transformation procedures known to the skilled man. Transformed seeds are selected and harvested and (where possible) transformed lines are established by normal selfing. Parallel transformations with 35S promoter constructs and the entire SERK-interacting gene itself are used as controls to evaluate over-expression in many cells or only in the few cells that naturally express the gene. The 35S promoter construct may give embryo formation wherever the signal that activates SERK-mediated transduction is present in the plant. A testing system based on emasculation and the generation of donor plant lines for pollen carrying LTP1 promoter-GUS and SERK promoter-barnase is established.

The same constructs (35S, EP3-1, AtChitIV, AtLTP-1 and SERK promoters fused to SERK-interacting coding sequences) can be employed for transformation into several *Arabidopsis* backgrounds such as wild type, male sterile, *fis* (allelic to *emb 173*) and *primordia timing (pt)-1* lines, or a combination of two or several of these backgrounds. The wt lines are used as a control to evaluate possible effects on normal zygotic embryogenesis, and to score for seed set without fertilization after emasculation. The *ms* lines are used to score directly for seed set without fertilization. The *fis* lines exhibit a certain degree of seed and embryo development without fertilization, so may be expected to have a natural tendency for apomictic embryogenesis, which may be enhanced by the presence of the constructs. The *pt-1* line has superior regenerative capabilities and has been used to initiate the first stably embryogenic *Arabidopsis* cell suspension cultures. Combinations of several of the above backgrounds are obtained by crossing with each other and with lines expressing SERK-interacting proteins ectopically. Except for the *ms* lines, propagation can proceed by normal selfing, and analysis of apomictic traits. A similar strategy is followed if the ATChitIV, AtLTP-1 and SERK promoters are replaced by the *bel-1* and *fbp-7* promoters as well by other promoters specific for components of the female gametophyte.

The invention still further includes vectors comprising DNA as indicated in the preceding paragraphs, plants transformed with the vector, progeny of such plants which contain the DNA stably incorporated, and the apomictic seeds of such plants or such progeny.

The genes to be expressed can be introduced into the plant cells in a number of art-recognized ways summarized in the paragraph bridging pages 7 and 8 of WO 97/43427.

Comprised within the scope of the present invention are transgenic plants, in particular transgenic fertile plants transformed by means of the aforescribed processes and their asexual and/or sexual progeny, which still contain the DNA stably incorporated, and/or the apomictic seeds of such plants or such progeny. Said plants can be used in the same way as described on pages 10 to 12 of WO 97/43427.

A transgenic plant according to the invention may be a dicotyledonous or a monocotyledonous plant. Such plants include field crops, vegetables and fruits including tomato, pepper, melon, lettuce, cauliflower, broccoli, cabbage, brussels sprout, sugar beet, corn, sweetcorn, onion, carrot, leek, cucumber, tobacco, alfalfa, aubergine, beet, broad bean, celery, chicory, cow pea, endive, gourd, groundnut, papaya, pea, peanut, pineapple, potato, safflower, snap bean, soybean, spinach, squashes, sunflower, sorghum, water-melon, and the like; and ornamental crops including Impatiens, Begonia, Petunia, Pelargonium, Viola, Cyclamen, Verbena, Vinca, Tagetes, Primula, Saint Paulia, Ageratum, Amaranthus, Anthirrhinum, Aquilegia, Chrysanthemum, Cineraria, Clover, Cosmo, Cowpea, Dahlia, Datura, Delphinium, Gerbera, Gladiolus, Gloxinia, Hippeastrum, Mesembryanthemum, Salpiglossis, Zinnia, and the like. In a preferred embodiment, the DNA is expressed in "seed crops" such as corn, sweet corn and peas etc. in such a way that the apomictic seed which results from such expression is not physically mutated or otherwise damaged in comparison with seed from untransformed like crops. Preferred are monocotyledonous plants of the *Graminaceae* family involving Lolium, Zea, Triticum, Triticale, Sorghum, Saccharum, Bromus, Oryzae, Avena, Hordeum, Secale and Setaria plants.

More preferred are transgenic maize, wheat, barley, sorghum, rye, oats, turf and forage grasses, millet, rice and sugar cane. Especially preferred are maize, wheat, sorghum, rye, oats, turf grasses and rice.

Among the dicotyledonous plants Arabidopsis, soybean, cotton, sugar beet, oilseed rape, tobacco and sunflower are more preferred herein. Especially preferred are tomato, pepper, melon lettuce, Brassica vegetables, soybean, cotton, tobacco, sugar beet and oilseed rape.

The expression 'progeny' is understood to embrace both, "asexually" and "sexually" generated progeny of transgenic plants. This definition is also meant to include all mutants and variants obtainable by means of known processes, such as for example cell fusion or mutant selection and which still exhibit the characteristic properties of the initial transformed plant, together with all crossing and fusion products of the transformed plant material. This also includes progeny plants that result from a backcrossing, as long as the said progeny plants still contain the DNA according to the invention.

Another object of the invention concerns proliferation material of the transgenic plants. It is defined relative to the invention as any plant material that may be propagated sexually or asexually *in vivo* or *in vitro*. Particularly preferred within the scope of the present invention are protoplasts, cells, calli, tissues, organs, seeds, embryos, pollen, egg cells, zygotes, together with any other propagating material obtained from transgenic plants.

Parts of plants, such as for example flowers, stems, fruits, leaves, roots originating in transgenic plants or their progeny previously transformed by means of the process of the invention and therefore consisting at least in part of transgenic cells, are also an object of the present invention. Especially preferred are apomictic seeds.

The present invention is exemplified by transgenic expression of a SERK-interacting gene in *Arabidopsis* under the control of plant expression signals, particularly a promoter which regulates expression of SERK genes *in planta*, but preferably a developmentally regulated or inducible promoter such as, for example, the carrot chitinase DcEP3-1 gene promoter, the *Arabidopsis* AtChitIV gene promoter, the *Arabidopsis* LTP-1 gene promoter, the *Arabidopsis* bel-1 gene promoter, the petunia fbp-7 gene promoter, the *Arabidopsis* ANT gene promoter, or the promoter of the O126 gene from *Phalaenopsis*; the *Arabidopsis* AtDMC1 promoter, or the pTA7001 inducible promoter.

The promoters of the DcEP3-1 and the AtChit IV genes may be cloned and characterized by standard procedures. The desired coding sequence is cloned behind the DcEP3-1, the AtChit IV or the AtLTP-1 promoters and transformed into *Arabidopsis*. The ligation is performed in such a way that the promoter is operably linked to the sequence to be transcribed. This construct, which also contains known marker genes providing for selection of transformed material, is inserted into the T-DNA region of a binary vector such as pBIN19 and transformed into *Arabidopsis*. *Agrobacterium*-mediated transformation into *Arabidopsis* is performed by the

vacuum infiltration or root transformation procedures known to the skilled man. Transformed seeds are selected and harvested and (where possible) transformed lines are established by normal selfing. Parallel transformations with 35S promoter constructs and the entire SERK-interacting gene itself are used as controls to evaluate over-expression in many cells or only in the few cells that naturally express the gene. The 35S promoter construct may give embryo formation wherever the signal that activates SERK-mediated transduction is present in the plant. A testing system based on emasculation and the generation of donor plant lines for pollen carrying LTP1 promoter-GUS and SERK promoter-barrenase is established.

The same constructs (35S, EP3-1, AtChitIV, AtLTP-1 and SERK promoters fused to the SERK-interacting coding sequence) are employed for transformation into several *Arabidopsis* backgrounds. These backgrounds are wild type, male sterile, *fis* (allelic to *emb 173*) and primordia timing (*pt*)-1 lines, or a combination of two or several of these backgrounds. The wt lines are used as a control to evaluate possible effects on normal zygotic embryogenesis, and to score for seed set without fertilization after emasculation. The *ms* lines are used to score directly for seed set without fertilization. The *fis* lines exhibit a certain degree of seed and embryo development without fertilization, so may be expected to have a natural tendency for apomictic embryogenesis, which may be enhanced by the presence of the constructs. The *pt*-1 line has superior regenerative capabilities and has been used to initiate the first stably embryogenic *Arabidopsis* cell suspension cultures. Combinations of several of the above backgrounds are obtained by crossing with each other and with lines expressing SERK-interacting proteins ectopically. Except for the *ms* lines, propagation can proceed by normal selfing, and analysis of apomictic traits. A similar strategy is followed in which the ATChitIV, AtLTP-1 and SERK promoters are replaced by the *bel*-1 and *fbp*-7 promoters as well by other promoters specific for components of the female gametophyte.

Whilst the present invention has been particularly described by way of the production of apomictic seed by heterologous expression of a SERK-interacting gene in the nucellar region of the carpel, the skilled man will recognize that other genes, the products of which have a similar structure/function may likewise be expressed with similar results. Moreover, although the example illustrates apomictic seed production in *Arabidopsis*, the invention is, of course, not limited to the expression of apomictic seed-inducing genes solely in this plant. Moreover, the present disclosure also includes the possibility of expressing the inventive gene sequences in

transformed plant material in a constitutive, tissue non-specific manner, for example under transcriptional control of a CaMV35S or NOS promoter.

The skilled man who has the benefit of the present disclosure will also recognize that a SERK-interacting genes may be transformed into plant material which may be propagated and/or differentiated and used as an explant from which somatic embryos can be obtained. Expression of such sequences in the transformed tissue substantially increases the percentage of the cells in the tissue which are competent to form somatic embryos, in comparison with the number present in non-transformed like tissue.

The following examples illustrate the isolation and cloning of genes encoding SERK-interacting proteins and the production of apomictic seed by heterologous expression of said genes in the nucellar region of the carpel so that somatic embryos form which penetrate the embryo sac and are encapsulated by the seed as it develops.

EXAMPLES

Example 1: *Isolation of Arabidopsis genes encoding proteins interacting with the Arabidopsis SERK gene product*

Construction of a SERK bait plasmid

The cDNA sequence of Arabidopsis SERK clone AtSERKtot61 in pBluescript SK- is used as the DNA template to amplify by PCR the SERK open reading frame devoid of its N-terminal sequence using the oligonucleotide primers

V6 (5' -ATGCTTTGCATAACTTTGAGG-3'; SEQ ID NO: 17) and

T7 (5' -AATACGACTCACTATAG-3'; SEQ ID NO: 18).

The resulting PCR product is cloned into the vector pGEM-T (Promega). From the resulting plasmid an NcoI-NotI fragment is isolated and cloned into the NcoI-NotI sites of the yeast lexA two hybrid bait vector pEG202 SERK (Origene). Nucleotide sequence analysis is performed to confirm the correct orientation and sequence of the PCR product in the resulting SERK bait plasmid. Bait protein expression and activity is determined using along the protocols described in Current Protocols in Molecular Biology 1996, chapter 20, supplement 33, contributed by E.A. Golemis; J. Gyuris and R. Brent. The construct is shown

to possess transcriptional activity in yeast strain EGY48. Furthermore, repressor activity on a reporter gene shows correct nuclear localization of the SERK gene product. Yeast transformed with the SERK bait plasmid proves to be leucine heterotrophic, indicating that the construct is not resulting in autoactivation of the *lexA* selection screen. The tests demonstrate that the SERK bait construct is suitable for *lexA* two hybrid screening.

Screening of a *lexA* two hybrid library

Yeast strain EGY48 transformed with the LacZ reporter plasmid pSH18-34 (Origene) and the bait vector pEG202 SERK is transformed with the cDNA library vector pJG4-5 (Origene) according to the LiAc/PEG4000 procedure described in Current protocols in Molecular Biology 1996, chapter 20, supplement 33, contributed by E.A. Golemis; J. Gyuris and R. Brent. A cDNA library from *Arabidopsis thaliana* young silique tissue containing early globular stage embryos is obtained (provided by Prof. Gerd Jürgens, Tuebingen). The primary library contains approximately 2.000.000 cDNA clones and the average insert length is 1.4 kB (as calculated from 90 clones of which the insert length varies from 0.2 to 4.5 kB). 10% of the clones contain no insert. The library is amplified once in *E. coli* before screening for SERK protein interaction. Induction of the fusion proteins in pJG4-5 is by the application of galactose in the medium. Under non-inducing conditions, yeast cells are grown in glucose and do not express the pJG4-5 fusion proteins. 4.200.000 prey cDNA clones are transformed into the yeast strain containing the pEG202 SERK bait plasmid and the pSH18-34 reporter plasmid. Transformation efficiency is up to 270.000 colonies per microgram of vector DNA. The plasmid pJG4-5 contains the *TRP1* selectable marker, pSH18-34 has an *URA3* selectable marker and pEG202 contains a *HIS3* selectable marker. Growth of the transformed yeast cells is taking place in complete minimal (CM) medium supplemented with either 2% glucose or 2% galactose + raffinose (in the latter case the galactose-inducible promoter on the vector pJG4-5 is activated, resulting in expression of the cDNA library fusion proteins. Yeast strain EGY48 contains six *LexA* operators which direct transcription from the *LEU2* gene. When both the SERK fusion protein and the cDNA library fusion protein are expressed the *LexA* DNA-binding domain of the SERK fusion protein can interact with the activation domain of the library cDNA fusion protein to form an active *LexA* transcription factor which in turn allows to select for leucine autotrophic transformants. The LacZ reporter construct on the plasmid pSH18-34 contains one *LexA*

operator in a promoter context different from the *LEU2* gene. Xgal and the presence of an active LexA transcription complex also allows determination of LacZ activity.

Triple selection for all three plasmids is performed on GLU/CM-his-ura-trp 24cm/24cm plates with approximately 100.000 colonies per plate. A total of 4.200.000 yeast primary transformants are obtained. The colonies are scraped from the plates with a sterile glass slide, collected in two different A or B labeled 50 ml tubes and frozen at -80°C. In order to estimate the colony titer a sample is plated on GAL/RAF/CM -ura-his-trp-leu plates. After determining the titer, library screening is continued by plating approximately 1.000.000 colonies on 10cm/10cm plates each. A total of 36.000.000 colonies is plated on leu selection plates GAL/CM-his-ura-trp-leu (20 million from vial A and 16 million from vial B). Colonies are isolated when the diameter of the colonies is at least 1 mm. The numbers of isolated colonies from each day and vial are indicated in the table below:

2 days	3 days	4 days
15A	93A	27A
9B	81B	25B

All isolated colonies are replated on different plates for determination of LacZ activity and only those colonies are selected which fit to the described criteria for each medium:
Numbers of isolated colonies from each day and vial are indicated:

GAL/RAF/CM	-ura-his-trp-leu	growth yes
GLU/CM	-ura-his-trp-leu	growth no
GAL/RAF/CM	-ura-his-trp + Xgal	blue and growth yes
GLU/CM	-ura-his-trp + Xgal	not blue, growth yes

<12 hours	20 hours	28 hours	48 hours	72 hours
4A	17A	9A	11A	24A
2B	6B	5B	15B	24B

A total of approximately 250 colonies is growing on leucine selection plates and tested for lacZ activity. 107 of these colonies show blue staining as an indication for lacZ activity.

Colony PCR performed on these 107 colonies with primers around the cloning site of the prey vector pJG4-5 generates approximately 10 different groups of cDNA clones based on PCR size. Sau3A1 digestion of the PCR fragments makes a more detailed grouping of different classes of SERK-interacting candidate cDNA clones possible. Members of all different classes are used to isolate and to clone the prey plasmid into *E.coli* and to determine the nucleotide and predicted amino acid sequence. Prey plasmids are retransformed in yeast and tested for SERK-dependent activation of leu selection and lacZ activity. All classes of cDNA clones prove to display a SERK-dependent yeast LexA two hybrid interaction after retransformation experiments. All these clones represent intracellular or membrane-attached factors involved in the signalling pathway mediated by the SERK receptor kinase protein. A total of 8 different classes of SERK-interacting proteins is identified.

Example 2: ***Function of SERK-interacting proteins***

Four of the classes of proteins that show an interaction with SERK are members of the family of Squamosa-promoter Binding Protein (SBP) transcription factors (Klein et al, Mol. Gen Genet 250: 7-16, 1996). They are represented by the clones 3A35 (SEQ ID NO: 1 and SEQ ID NO: 2), 3B39 (SEQ ID NO: 3 and SEQ ID NO: 4), 4B19 (SEQ ID NO: 5 and SEQ ID NO: 6), and 3A52 (SEQ ID NO: 7 and SEQ ID NO: 8). These proteins are able to interact specifically with DNA through a conserved domain of 79 amino acid residues, the SBP-box. Within the SBP-box a remarkable arrangement of cysteine and histidine residues can be recognized, which is reminiscent of zinc-fingers and probably involved in the recognition of specific promoter elements. A bipartite nuclear localization signal is placed at the C-terminal end of the SBP-box (Dingwall et al, Trends Biochem Sci 16: 478-481, 1991). Both the N-terminal and the C-terminal domains of the SERK-interacting SBP proteins are highly variable and are probably involved in regulation of protein activity. One of the classes of SBP proteins, represented by 4B19, is identical with SPL3, a gene involved in the floral transition and expressed in developing flower buds (Cardon and Hohmann 1997 Plant Journal 12, 367-377). The most likely model for the signalling pathway mediated by the SERK and SBP proteins is transphosphorylation of cytoplasmic SBP-transcription factors by SERK after ligand binding, followed by nuclear translocation of the factors and binding to specific regulatory DNA target sites on the genome. A similar mode of signal transduction

has been described for animal serine-threonine receptor-kinase proteins which are known to transphosphorylate a family of so called SMAD transcription factors. Phosphorylated activated SMAD proteins are translocated into the nucleus (Heldin et al, Nature 390: 465-471, 1997).

Another class of SERK-interacting proteins is represented by an isoform of the family of 14-3-3 proteins. 4B11 (SEQ ID NO: 9 and SEQ ID NO: 10) is identical to the 14-3-3 type lambda protein (Wu et al, Plant Physiol 114: 1421-1431, 1997). A total of 10 different 14-3-3 proteins is present in *Arabidopsis* and the different members are involved in intracellular signal transduction. They mediate signal transduction by binding to phosphoserine-containing proteins on specific binding motifs represented by conserved amino acid sequences like RxxS(p)xP (Yaffe et al, Cell 91: 961-971, 1997). A putative 14-3-3 interaction domain having the amino acid sequence RPPSQP is also found at position 391-396 of the *Arabidopsis* SERK protein, and at the corresponding aligned region of the *Daucus carota* SERK protein having the amino acid sequence RQPSEP providing SERK with a mechanism for a 14-3-3 mediated signal transduction.

4A24 (SEQ ID NO: 11 and SEQ ID NO: 12) represents a member of a small new *Arabidopsis* gene family from which one member has already been described in the literature as the NDR1 protein (Century et al, Science 278: 1963-1965, 1997). NDR1 is likely to encode a membrane-associated component in the signal transduction pathway downstream of pathogen-recognizing proteins. It was suggested that NDR1 is a protein that interacts with many different receptors to transduce their signal. 4A24 represents a new member in this small family of proteins and might have an important function in intracellular signal transduction mediated by transmembrane receptors.

Clone 3B76 (SEQ ID NO: 13 and SEQ ID NO: 14) encodes a protein with homology to a domain in *E.coli* aminopeptidase N. and might encode an *Arabidopsis* protease, interacting or activated by SERK.

The predicted amino acid sequence represented by clone 4A5 (SEQ ID NO: 15 and SEQ ID NO: 16) has no homology with known gene products although there is a small not yet described family of related gene products in *Arabidopsis* (AA585806, AA651106, T45539).

Example 3: *Transformation of Arabidopsis with genes encoding SERK-interacting proteins*

Plasmids containing promoter sequences

- The CaMV 35S promoter enhanced by duplication of the -343 to -90 region (Kay et al, Science 236: 1299-1302, 1987) is isolated from the mMON999 vector by digestion with HindIII and SstI and cloned into the pBluescript SK- vector resulting in vector pMT120.
- The promoter of the FBP7 gene from *Petunia* (Angenent et al, Plant Cell 7: 1569-1582, 1995) is cloned by subcloning the 0.6 kb HindIII-XbaI genomic DNA fragment of FBP7 into the HindIII-XbaI site of pBluescript KS- resulting in the vector FBP201.

Plasmids containing full length SERK-interacting cDNA clones

Full length cDNA of the identified SERK-interacting gene products is produced by RT-PCR amplification of early stage *Arabidopsis* silique RNA. Full length cDNA is isolated from clones 3A35, 3A52, and 4B19. Clone 3B39 was already present as a full length cDNA clone. Oligo sequences are based on the nucleotide sequences from identical BAC or EST clones.

Binary vector constructs

Based on the pBIN19 vector, a binary vector is constructed for transformation of the *Arabidopsis thaliana* SERK-interacting cDNA under the control of different promoters. The full length cDNA clones of the putative SBP-transcription factors interacting with SERK are blunted by Klenow treatment and cloned into the SmaI site of pBIN19. The polyadenylation sequence from the pea *rbcS::E9* gene (Millar et al, Plant Cell 4: 1075-1087, 1992) is placed downstream from the coding sequence by cloning a Klenow-filled EcoRI-HindIII E9 DNA fragment into the Klenow-filled XmaI site of the pBIN19:SERK interacting factor in order to generate the binary vectors pAt3A35, pAt3A52, pAt4B19 and pAt3B39. The pAt binary vectors are used to generate promoter-SERK interacting factor constructs.

- The CaMV 35S promoter is cloned in the SmaI site of the pAt vector constructs as a Klenow-filled KpnI-SstI fragment to give p35SA vectors.
- The SacI-KpnI fragment of FBP201 is filled with Klenow and cloned into the SmaI site of the pAt vector constructs to give the pFBP201At vectors.

Introduction of plant expression vectors into *Arabidopsis thaliana* plants

The above described vector constructs are electrotransformed into *Agrobacterium tumefaciens* strain C58C1. Wild type *Arabidopsis thaliana* WS plants are grown under standard long day conditions (16 hours light and 8 hours dark). The first emerging inflorescence is removed in order to increase the number of inflorescences. Five days later, plants are used for the vacuum infiltration procedure. Transformed *Agrobacterium* C58C1 is grown on LB plates with 50 mg/l kanamycin, 50 mg/l rifampicin and 25 mg/l gentamycin. Single colonies are used to inoculate 500 ml of liquid medium (as described above) and grown O/N at 28°C. Log phase culture ($OD_{600}=0.8$) is centrifuged to pellet cells and resuspended in 150 ml of infiltration medium (0.5 x MS medium pH 5.7, 5% sucrose and 1 mg/l benzylaminopurine). The inflorescences of 6 *Arabidopsis* plants are submerged in the infiltration suspension while the remaining parts of the plants (which are still potted) are placed upside down on meshed wire to avoid contact with the infiltration medium. Vacuum is applied to the whole set-up for 10 min at 50 kPa. Plants are directly afterwards placed under standard long day conditions. After completed seed setting the seeds are surface sterilized by an 1% sodium hypochlorite soak, thoroughly washed with sterile water and planted onto petridishes with 0.5 x MS medium, 1% agar and 80 mg/l kanamycin in order to select for transformed seeds. After 7 days of germination under long day conditions (10.000 lux) the transformed seedlings can be identified by their green colour of their cotyledons and the appearance of the first true leaves. Transformed seedlings are further grown in soil under long day conditions. The vacuum infiltration method results in approximately 0.1% transformed seeds.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:

- (A) NAME: NOVARTIS AG
- (B) STREET: Schwarzwaldallee 215
- (C) CITY: Basel
- (E) COUNTRY: Switzerland
- (F) POSTAL CODE (ZIP): 4058
- (G) TELEPHONE: +41 61 324 11 11
- (H) TELEFAX: + 41 61 322 75 32

(ii) TITLE OF INVENTION: Organic Compounds

(iii) NUMBER OF SEQUENCES: 18

(iv) COMPUTER READABLE FORM:

- (A) MEDIUM TYPE: Floppy disk
- (B) COMPUTER: IBM PC compatible
- (C) OPERATING SYSTEM: PC-DOS/MS-DOS
- (D) SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 551 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Arabidopsis thaliana

(vii) IMMEDIATE SOURCE:

- (B) CLONE: 3A35

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

ACGTGTCCGT GGAGGCGGGT CGGGTCAGTC GGGTCAGATA CCAAGGTGCC AAGTGAAGG	60
TTGTGGGATG GATCTAACCA ATGCAAAAGG TTATTACTCG AGACACCGAG TTTGTGGAGT	120
GCACTCTAAA ACACCTAAAG TCACTGTGGC TGGTATCGAA CAGAGGTTTT GTCAACAGTG	180
CAGCAGGTTT CATCAGCTTC CGGAATTTGA CCTAGAGAAA AGGAGTTGCC GCAGGAGACT	240

CGCTGGTCAT AATGAGCGAC GAAGGAAGCC ACAGCCTGCG TCTCTCTCTG TGTTAGCTTC 300
 TCGTTACGGG AGGATCGCAC CTTGCTTTTA CGAAAATGGT GATGCTGGAA TGAATGGAAG 360
 CTTTCTTGGG AACCAAGAGA TAGGATGGCC AAGTTCAAGA ACATTGGATA CAAGAGTGAT 420
 GAGGCGGCCA GTGTCATCAC CGTCATGGCA GATCAATCCA ATGAATGTAT TTAGTCAAGG 480
 TTCAGTTGGT GGAGGAAGGA CAAGCTTCTC ATCTCCAGAG ATTATGGACA CTAAACTAGA 540
 GAGCTACAAG G 551

(2) INFORMATION FOR SEQ ID NO: 2:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 375 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (iii) HYPOTHETICAL: NO
- (iii) ANTI-SENSE: NO
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Arabidopsis thaliana
- (vii) IMMEDIATE SOURCE:
 - (B) CLONE: 3A35
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

Met Glu Met Gly Ser Asn Ser Gly Pro Gly His Gly Pro Gly Gln Ala
 1 5 10 15

Glu Ser Gly Gly Ser Ser Thr Glu Ser Ser Ser Phe Ser Gly Gly Leu
 20 25 30

Met Phe Gly Gln Lys Ile Tyr Phe Glu Asp Gly Gly Gly Gly Ser Gly
 35 40 45

Ser Ser Ser Ser Gly Gly Arg Ser Asn Arg Arg Val Arg Gly Gly Gly
 50 55 60

Ser Gly Gln Ser Gly Gln Ile Pro Arg Cys Gln Val Glu Gly Cys Gly
 65 70 75 80

Met Asp Leu Thr Asn Ala Lys Gly Tyr Tyr Ser Arg His Arg Val Cys
 85 90 95

Gly Val His Ser Lys Thr Pro Lys Val Thr Val Ala Gly Ile Glu Gln

- 23 -

100	105	110
Arg Phe Cys Gln Gln Cys Ser Arg Phe His Gln Leu Pro Glu Phe Asp		
115	120	125
Leu Glu Lys Arg Ser Cys Arg Arg Arg Leu Ala Gly His Asn Glu Arg		
130	135	140
Arg Arg Lys Pro Gln Pro Ala Ser Leu Ser Val Leu Ala Ser Arg Tyr		
145	150	155
Gly Arg Ile Ala Pro Ser Leu Tyr Glu Asn Gly Asp Ala Gly Met Asn		
165	170	175
Gly Ser Phe Leu Gly Asn Gln Glu Ile Gly Trp Pro Ser Ser Arg Thr		
180	185	190
Leu Asp Thr Arg Val Met Arg Arg Pro Val Ser Ser Pro Ser Trp Gln		
195	200	205
Ile Asn Pro Met Asn Val Phe Ser Gln Gly Ser Val Gly Gly Gly Arg		
210	215	220
Thr Ser Phe Ser Ser Pro Glu Ile Met Asp Thr Lys Leu Glu Ser Tyr		
225	230	235
Lys Gly Ile Gly Asp Ser Asn Cys Ala Leu Ser Leu Leu Ser Asn Pro		
245	250	255
His Gln Pro His Asp Asn Asn Asn Asn Asn Asn Asn Asn Asn Asn		
260	265	270
Asn Asn Asn Thr Trp Arg Ala Ser Ser Gly Phe Gly Pro Met Thr Val		
275	280	285
Thr Met Ala Gln Pro Pro Pro Ala Pro Ser Gln His Gln Tyr Leu Asn		
290	295	300
Pro Pro Trp Val Phe Lys Asp Asn Asp Asn Asp Met Ser Pro Val Leu		
305	310	315
Asn Leu Gly Arg Tyr Thr Glu Pro Asp Asn Cys Gln Ile Ser Ser Gly		
325	330	335
Thr Ala Met Gly Glu Phe Glu Leu Ser Asp His His His Gln Ser Arg		
340	345	350
Arg Gln Tyr Met Glu Asp Glu Asn Thr Arg Ala Tyr Asp Ser Ser Ser		
355	360	365
His His Thr Asn Trp Ser Leu		
370	375	

(2) INFORMATION FOR SEQ ID NO: 3:

- 24 -

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 859 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA to mRNA
- (iii) HYPOTHETICAL: NO
- (iii) ANTI-SENSE: NO
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: *Arabidopsis thaliana*
- (vii) IMMEDIATE SOURCE:
 - (B) CLONE: 3B39

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

TCAACATTGC TTCCTAACCA GAAATCCACC ATCATCTTCC CACGAATACA ACTTAAAGCT	60
TTACCAGAAA ATGGAGGGTC AGAGAACACA ACGCCGGGGT TACTTGAAAG ACAAGGCTAC	120
AGTCTCCAAC CTTGTTGAAG AAGAAATGGA GAATGGCATG GATGGAGAAG AGGAGGATGG	180
AGGAGACGAA GACAAAAGGA AGAAGGTGAT GGAAAGAGTT AGAGGTCTTA GCACTGACCG	240
TGTTCCATCG CGACTGTGCC AGGTCGATAG GTGCACTGTT AATTGACTG AGGCCAAGCA	300
GTATTACCGC AGACACAGAG TATGTGAAGT ACATGCAAAG GCATCTGCTG CGACTGTTGC	360
AGGGGTCAGG CAACGCTTTT GTCAACAATG CAGCAGGTTT CATGAGCTAC CAGAGTTTGA	420
TGAAGCTAAA AGAAGCTGCA GGAGGCGCTT AGCTGGACAC AATGAGAGGA GGAGGAAGAT	480
CTCTGGTGAC AGTTTTGGAG AAGGGTCAGG CCGGAGAGGG TTTAGCGGTC AACTGATCCA	540
GACTCAAGAA AGAAACAGGG TAGACAGGAA ACTTCCTATG ACCAACTCAT CATTTAAGGG	600
ACCACAGATC AGATAAACC TCCCGCTCTC TCTCTCTGT CATCTACATA TGCTCTATCT	660
ACACTCTTAT TAGACAAATA ATGGCATCTA ACAATGTCAA GAAAAGTTGG TCATGGTATT	720
AAATCCTAGA GGGAAATATA AGTATAAACC TTTAGTCCCC TTTATGCTGT CCTGTAATGA	780
ATATCTATCC GGAAATGTAT TCGCATAGTC TTGCGTCTAA TAATGTTTAT TAAAAAAAAA	840
AAAAAAAAAA AAAAAAAAAA	859

(2) INFORMATION FOR SEQ ID NO: 4:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 181 amino acids

(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

```
(ii) MOLECULE TYPE: protein
(iii) HYPOTHETICAL: NO
(iii) ANTI-SENSE: NO
(vi) ORIGINAL SOURCE:
      (A) ORGANISM: Arabidopsis thaliana
(vii) IMMEDIATE SOURCE:
      (B) CLONE: 3B39
```

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

[illegible]

(2) INFORMATION FOR SEQ ID NO: 5:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 479 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA to mRNA
- (iii) HYPOTHETICAL: NO
- (iii) ANTI-SENSE: NO
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Arabidopsis thaliana
- (vii) IMMEDIATE SOURCE:
 - (B) CLONE: 4B19

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

```

AGAAGCAGAA AGGTAAAGCT ACAAGTAGTA GTGGAGTTTG TCAGGTCGAG AGTTGTACCG      60
CGGATATGAG CAAAGCCAAA CAGTACCACA AACGACACAA AGTCTGCCAG TTTCATGCCA      120
AAGCTCCTCA TGTTCGGATC TCTGGTCTTC ACCAACGTTT CTGCCAACAA TGCAGCAGGT      180
TTCACGCGCT CAGTGAGTTT GATGAAGCCA AGCGGAGTTG CAGGAGACGC TTAGCTGGAC      240
ACAACGAGAG AAGGCGGAAA AGCACAAC TG ACTAAAGACG GTGAAACGTG TGAGATCCCG      300
GTTTGAAGGT TAATGAAACA GGCTTTGCTT ACTCTCTTCT GTCAGTCTCT TTTAGCTCCT      360
TGTAATCCTC TGTGTCTCTG TCTGTTTCTC CATATTACCT GTAATCAAAG CTATCTGCTA      420
AACCTACGAC ATGGTTAAAT AAATGCATIG AGACTTAAAA AAAAAAAAAA AAAAAAAAAA      479

```

(2) INFORMATION FOR SEQ ID NO: 6:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 131 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (iii) HYPOTHETICAL: NO
- (iii) ANTI-SENSE: NO
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Arabidopsis thaliana

(vii) IMMEDIATE SOURCE:
(B) CLONE: 4B19.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

[illegible]

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2682 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: *Arabidopsis thaliana*

(vii) IMMEDIATE SOURCE:

(B) CLONE: 3A52

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

GCCATTCAAG GAGACACTAA TGGTGCTCTT ACTTTGAATC TTAATGGTGA AAGTGATGGC	60
CTTTTTCCTG CCAAGAAGAC CAAATCCGGA GCCGTTTGTC AGGTGAAAA CTGTGAAGCT	120
GATCTTAGTA AAGTTAAGGA TTATCATAGA CGCCATAAGG TCTGTGAGAT GCATTCCAAG	180
GCTACTAGTG CCACTGTCGG AGGTATCTTG CAGCGCTTTT GTCAGCAATG TAGTAGGTTT	240
CATCTTCTGC CAGGTTTCGA TGACGGAAAG AGAAGTTGTC GTAGACGTTT GGCTGGCCAT	300
AATAAACGTC CGAGGAAAAC AAATCCCGAA CCTGGCGCTA ACGGGAATCC TAGTGATGAT	360
CACTCAAGCA ACTATCTCTT GATTACTCTC TTGAAGATAC TCTCCAATAT GCATAACCAT	420
ACCGGTGATC AAGATTTGAT GTCTCATCTT CTGAAGAGCC TCGTAAGCCA TGCTGGCGAA	480
CAGTTAGGGA AAAACTTAGT TGAACCTCTT CTACAAGGAG AGATCTCAAG GTTCCTTAAA	540
ATATTGGAAA ACTCGGCTTT GCTTGGGATT GAGCAAGCTC CTCAAGAGGA GTTAAAGCAA	600
TTTTCGGCTC GGCAAGATGG GACAGCTACC GAGAACAGAT CAGAAAAACA AGTCAAAATG	660
AATGATTTTG ATTTGAATGA TATCTATATA GACTCAGATG ACACAGACGT CGAAAGATCT	720
CCTCCTCCAA CGAATCCAGC GACCAGTTCT CTTGATTATC CTTTCATGGAT ACATCAGTCT	780
AGTCCGCCTC AGACAAGTAG GAATTCAGAT TCAGCATCTG ACCAGTCACC CTCAAGTTCT	840
AGTGAAGATG CTCAGATGCG CACAGGCCGG ATTGTGTTCA AACTATTTGG GAAAGAGCCA	900
AATGAATTTT CTATTGTCTT ACGAGGACAG ATTCTTGACT GGTTATCGCA TAGTCCAAC	960
GACATGGAGA GCTACATAAG ACCTGGCTGT ATCGTATTGA CCATCTATCT TCGTCAAGCT	1020
GAAACTGCTT GGAAGAAGT TTCAGACGAT CTGGGTTTTA GCTTAGGGAA GCTTCTAGAT	1080
CTCTCCGATG ATCCCTTG TGACAACTGA TGGATTATG TAGGGTGCAG AACCAACTTG	1140
CATTGTGATA TAACGGTCAG GTTGTCGTTG ACACTTCATT GTCTCTAAAA AGTCGTGATT	1200
ATAGTCACAT CATTAGCGTT AAACCGCTTG CTATAGCTGC AACGGAGAAG GCTCAATTTA	1260
CAGTTAAAGG CATGAATCTC CGTCGGCGTG GCACAAGGTT ACTTTGTTCT GTTGAAGGAA	1320
AATACTTGAT TCAGGAAACA ACACACGATT CGACGACCAG GGAGGATGAC GATTTCAAGG	1380
ACAACAGTGA GATTGTTGAG TGTGTAAACT TCTCTTGTA TATGCCTATA TTGAGTGGTC	1440
GAGGATTCAT GGAGATTGAA GACCAAGGAC TCAGTAGCAG CTTCTTCCCT TTCTTAGTGG	1500
TTGAAGATGA CGATGTTTGT TCTGAAATCC GTATACTTGA AACCACATTA GAGTTCACTG	1560
GAACGATTTC TGCTAAGCAA GCTATGGATT TCATACATGA AATCGGTTGG CTTCTTCACA	1620

GAAGTAAACT TGGGGAATCA GACCCAAATC CAGGCGTTTT CCCATTAATA CGCTTCCAGT	1680
GGCTAATCGA GTTCTCAATG GATCGAGAGT GGTGCGCTGT GATCAGAAAG CTATTAAACA	1740
TGTTCTTTGA TGGAGCTGTT GGTGAATTTT CTTCTCTCTC TAATGCCACA CTGTCAGAAC	1800
TGTGCCTTCT TCACAGAGCC GTGAGGAAAA ACTCTAAGCC TATGGTTGAA ATGCTCTTGA	1860
GATATATTCC CAAGCAACAG AGAAACAGCT TGTTTAGACC CGATGCTGCT GGTCCAGCCG	1920
GCTTAACACC TCTTCATATT GCAGCTGGTA AAGACGGTTC AGAAGATGTG TTGGATGCGC	1980
TAACAGAAGA TCCTGCAATG GTGGGGATTG AAGCGTGGAA GACATGTCGA GACAGCACAG	2040
GCTTCACACC AGAAGACTAC GCACGCTTAC GCGGTCAC TTTCATACATC CACTTGATTG	2100
AACGCAAGAT CAATAAAAAG TCAACAAC TGATCATGTG TGTGGTCAAC ATCCCAGTTT	2160
CTTTCTCAGA CAGAGAGCAG AAAGAACCAA AATCAGGTCC GATGGCTTCA GCCTTGGAGA	2220
TCACACAGAT TCCATGCAAG CTCTGTGACC ATAAACTGGT GTATGGGACA ACACGCAGGT	2280
CTGTAGCGTA CAGACCAGCT ATGTTGTCAA TGGTGGCGAT TGCTGCGGTT TCGCTCTGTG	2340
TGGCACTTCT GTTTAAGAGT TGCCCGGAAG TGCTCTATGT GTTTCGAACCG TTCAGGTGGG	2400
AGTTATTGGA CTATGGAACA AGCTGAGTGT AAGTCTACTT TGAAAGATCT TCTAAGATAT	2460
ATATATGAAT GTTACTTATA TAAAACCATA GAGGTGTGAT TTCTATATGT AACTATATGA	2520
GTATAAGATA TAGAGACATG TTGGAGAAGA AGATTGTTGT TATTATTGTT GTTGTGTTTG	2580
TTGTGTAAAA GCCTCTCCTA TCTCTCTCGA ACCTAAGGAT TCTCTCTCTG ATTAGTATAT	2640
TTTTTGTTTG AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AA	2682

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 848 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: *Arabidopsis thaliana*

(vii) IMMEDIATE SOURCE:

- 30 -

(B) CLONE: 3A52

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

Met	Glu	Ala	Arg	Ile	Asp	Glu	Gly	Gly	Glu	Ala	Gln	Gln	Phe	Tyr	Gly	1	5	10	15
Ser	Val	Gly	Asn	Ser	Ser	Asn	Ser	Ser	Ser	Ser	Cys	Ser	Asp	Glu	Gly	20	25	30	
Asn	Asp	Lys	Lys	Arg	Arg	Ala	Val	Ala	Ile	Gln	Gly	Asp	Thr	Asn	Gly	35	40	45	
Ala	Leu	Thr	Leu	Asn	Leu	Asn	Gly	Glu	Ser	Asp	Gly	Leu	Phe	Pro	Ala	50	55	60	
Lys	Lys	Thr	Lys	Ser	Gly	Ala	Val	Cys	Gln	Val	Glu	Asn	Cys	Glu	Ala	65	70	75	80
Asp	Leu	Ser	Lys	Val	Lys	Asp	Tyr	His	Arg	Arg	His	Lys	Val	Cys	Glu	85	90	95	
Met	His	Ser	Lys	Ala	Thr	Ser	Ala	Thr	Val	Gly	Gly	Ile	Leu	Gln	Arg	100	105	110	
Phe	Cys	Gln	Gln	Cys	Ser	Arg	Phe	His	Leu	Leu	Pro	Gly	Phe	Asp	Asp	115	120	125	
Gly	Lys	Arg	Ser	Cys	Arg	Arg	Arg	Leu	Ala	Gly	His	Asn	Lys	Arg	Pro	130	135	140	
Arg	Lys	Thr	Asn	Pro	Glu	Pro	Gly	Ala	Asn	Gly	Asn	Pro	Ser	Asp	Asp	145	150	155	160
His	Ser	Ser	Asn	Tyr	Leu	Leu	Ile	Thr	Leu	Leu	Lys	Ile	Leu	Ser	Asn	165	170	175	
Met	His	Asn	His	Thr	Gly	Asp	Gln	Asp	Leu	Met	Ser	His	Leu	Leu	Lys	180	185	190	
Ser	Leu	Val	Ser	His	Ala	Gly	Glu	Gln	Leu	Gly	Lys	Asn	Leu	Val	Glu	195	200	205	
Leu	Leu	Leu	Gln	Gly	Arg	Arg	Ser	Gln	Gly	Ser	Leu	Asn	Ile	Gly	Asn	210	215	220	
Ser	Ala	Leu	Leu	Gly	Ile	Glu	Gln	Ala	Pro	Gln	Glu	Glu	Leu	Lys	Gln	225	230	235	240
Phe	Ser	Ala	Arg	Gln	Asp	Gly	Thr	Ala	Thr	Glu	Asn	Arg	Ser	Glu	Lys	245	250	255	
Gln	Val	Lys	Met	Asn	Asp	Phe	Asp	Leu	Asn	Asp	Ile	Tyr	Ile	Asp	Ser	260	265	270	

- 31 -

Asp Asp Thr Asp Val Glu Arg Ser Pro Pro Pro Thr Asn Pro Ala Thr
 275 280 285
 Ser Ser Leu Asp Tyr Pro Ser Trp Ile His Gln Ser Ser Pro Pro Gln
 290 295 300
 Thr Ser Arg Asn Ser Asp Ser Ala Ser Asp Gln Ser Pro Ser Ser Ser
 305 310 315 320
 Ser Glu Asp Ala Gln Met Arg Thr Gly Arg Ile Val Phe Lys Leu Phe
 325 330 335
 Gly Lys Glu Pro Asn Glu Phe Pro Ile Val Leu Arg Gly Gln Ile Leu
 340 345 350
 Asp Trp Leu Ser His Ser Pro Thr Asp Met Glu Ser Tyr Ile Arg Pro
 355 360 365
 Gly Cys Ile Val Leu Thr Ile Tyr Leu Arg Gln Ala Glu Thr Ala Trp
 370 375 380
 Glu Glu Leu Ser Asp Asp Leu Gly Phe Ser Leu Gly Lys Leu Leu Asp
 385 390 395 400
 Leu Ser Asp Asp Pro Leu Trp Thr Thr Gly Trp Ile Tyr Val Arg Val
 405 410 415
 Gln Asn Gln Leu Ala Phe Val Tyr Asn Gly Gln Val Val Val Asp Thr
 420 425 430
 Ser Leu Ser Leu Lys Ser Arg Asp Tyr Ser His Ile Ile Ser Val Lys
 435 440 445
 Pro Leu Ala Ile Ala Ala Thr Glu Lys Ala Gln Phe Thr Val Lys Gly
 450 455 460
 Met Asn Leu Arg Arg Arg Gly Thr Arg Leu Leu Cys Ser Val Glu Gly
 465 470 475 480
 Lys Tyr Leu Ile Gln Glu Thr Thr His Asp Ser Thr Thr Arg Glu Asp
 485 490 495
 Asp Asp Phe Lys Asp Asn Ser Glu Ile Val Glu Cys Val Asn Phe Ser
 500 505 510
 Cys Asp Met Pro Ile Leu Ser Gly Arg Gly Phe Met Glu Ile Glu Asp
 515 520 525
 Gln Gly Leu Ser Ser Ser Phe Phe Pro Phe Leu Val Val Glu Asp Asp
 530 535 540
 Asp Val Cys Ser Glu Ile Arg Ile Leu Glu Thr Thr Leu Glu Phe Thr
 545 550 555 560

- 32 -

Gly	Thr	Asp	Ser	Ala	Lys	Gln	Ala	Met	Asp	Phe	Ile	His	Glu	Ile	Gly	565	570	575
Trp	Leu	Leu	His	Arg	Ser	Lys	Leu	Gly	Glu	Ser	Asp	Pro	Asn	Pro	Gly	580	585	590
Val	Phe	Pro	Leu	Ile	Arg	Phe	Gln	Trp	Leu	Ile	Glu	Phe	Ser	Met	Asp	595	600	605
Arg	Glu	Trp	Cys	Ala	Val	Ile	Arg	Lys	Leu	Leu	Asn	Met	Phe	Phe	Asp	610	615	620
Gly	Ala	Val	Gly	Glu	Phe	Ser	Ser	Ser	Ser	Asn	Ala	Thr	Leu	Ser	Glu	625	630	635
Leu	Cys	Leu	Leu	His	Arg	Ala	Val	Arg	Lys	Asn	Ser	Lys	Pro	Met	Val	645	650	655
Glu	Met	Leu	Leu	Arg	Tyr	Ile	Pro	Lys	Gln	Gln	Arg	Asn	Ser	Leu	Phe	660	665	670
Arg	Pro	Asp	Ala	Ala	Gly	Pro	Ala	Gly	Leu	Thr	Pro	Leu	His	Ile	Ala	675	680	685
Ala	Gly	Lys	Asp	Gly	Ser	Glu	Asp	Val	Leu	Asp	Ala	Leu	Thr	Glu	Asp	690	695	700
Pro	Ala	Met	Val	Gly	Ile	Glu	Ala	Trp	Lys	Thr	Cys	Arg	Asp	Ser	Thr	705	710	715
Gly	Phe	Thr	Pro	Glu	Asp	Tyr	Ala	Arg	Leu	Arg	Gly	His	Phe	Ser	Tyr	725	730	735
Ile	His	Leu	Ile	Gln	Arg	Lys	Ile	Asn	Lys	Lys	Ser	Thr	Thr	Glu	Asp	740	745	750
His	Val	Val	Val	Asn	Ile	Pro	Val	Ser	Phe	Ser	Asp	Arg	Glu	Gln	Lys	755	760	765
Glu	Pro	Lys	Ser	Gly	Pro	Met	Ala	Ser	Ala	Leu	Glu	Ile	Thr	Gln	Ile	770	775	780
Pro	Cys	Lys	Leu	Cys	Asp	His	Lys	Leu	Val	Tyr	Gly	Thr	Thr	Arg	Arg	785	790	795
Ser	Val	Ala	Tyr	Arg	Pro	Ala	Met	Leu	Ser	Met	Val	Ala	Ile	Ala	Ala	805	810	815
Val	Cys	Val	Cys	Val	Ala	Leu	Leu	Phe	Lys	Ser	Cys	Pro	Glu	Val	Leu	820	825	830
Tyr	Val	Phe	Gln	Pro	Phe	Arg	Trp	Glu	Leu	Leu	Asp	Tyr	Gly	Thr	Ser	835	840	845

- 33 -

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 576 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: *Arabidopsis thaliana*

(vii) IMMEDIATE SOURCE:

(B) CLONE: 4B11

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

CAGCGGAAGA GCTCACCGTT GAAGAGAGGA ATCTCCTCTC TGTTGCTTAC AAAAACGTGA	60
TCGGATCTCT ACGCGCCGCC TGGAGGATCG TGTCTTCGAT TGAGCAGAAG GAAGAGAGTA	120
GGAAGAACGA CGAGCACGTG TCGCTTGTC AAGATTACAG ATCTAAAGTT GAGTCTGAGC	180
TTTCTTCTGT TTGCTCTGGA ATCCTTAAGC TCCTTGACTC GCATCTGATC CCATCTGCTG	240
GAGCGAGTGA GTCTAAGGTC TTTTACTTGA AGATGAAAGG TGATTATCAT CGGTACATGG	300
CTGAGTTTAA GTCTGGTGAT GAGAGGAAAA CTGCTGCTGA AGATACCATG CTCGCTTACA	360
AAGCAGCTCA GGATATCGCA GCTGCGGATA TGGCACCTAC TCATCCGATA AGGCTTGGTC	420
TGGCCCTGAA TTTCTCAGTG TTCTACTATG AGATTCTCAA TTCTTCAGAC AAAGCTTGTA	480
ACATGGCCAA ACAGGCTTTT GAGGAAGCCA TAGCTGAGCT TGACACTCTG GGAGAAGAAT	540
CCTACAAAGA CAGCACTCTC ATAATGCAGT TGCTGA	576

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 248 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

- 34 -

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: *Arabidopsis thaliana*

(vii) IMMEDIATE SOURCE:

(B) CLONE: 4B11

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

```

Met Ala Ala Thr Leu Gly Arg Asp Gln Tyr Val Tyr Met Ala Lys Leu
1           5           10           15

Ala Glu Gln Ala Glu Arg Tyr Glu Glu Met Val Gln Phe Met Glu Gln
          20           25           30

Leu Val Thr Gly Ala Thr Pro Ala Glu Glu Leu Thr Val Glu Glu Arg
      35           40           45

Asn Leu Leu Ser Val Ala Tyr Lys Asn Val Ile Gly Ser Leu Arg Ala
      50           55           60

Ala Trp Arg Ile Val Ser Ser Ile Glu Gln Lys Glu Glu Ser Arg Lys
65           70           75           80

Asn Asp Glu His Val Ser Leu Val Lys Asp Tyr Arg Ser Lys Val Glu
          85           90           95

Ser Glu Leu Ser Ser Val Cys Ser Gly Ile Leu Lys Leu Leu Asp Ser
      100           105           110

His Leu Ile Pro Ser Ala Gly Ala Ser Glu Ser Lys Val Phe Tyr Leu
      115           120           125

Lys Met Lys Gly Asp Tyr His Arg Tyr Met Ala Glu Phe Lys Ser Gly
      130           135           140

Asp Glu Arg Lys Thr Ala Ala Glu Asp Thr Met Leu Ala Tyr Lys Ala
      145           150           155           160

Ala Gln Asp Ile Ala Ala Ala Asp Met Ala Pro Thr His Pro Ile Arg
          165           170           175

Leu Gly Leu Ala Leu Asn Phe Ser Val Phe Tyr Tyr Glu Ile Leu Asn
      180           185           190

Ser Ser Asp Lys Ala Cys Asn Met Ala Lys Gln Ala Phe Glu Glu Ala
      195           200           205

Ile Ala Glu Leu Asp Thr Leu Gly Glu Glu Ser Tyr Lys Asp Ser Thr
      210           215           220

Leu Ile Met Gln Leu Leu Arg Asp Asn Leu Thr Leu Trp Thr Ser Asp
      225           230           235           240

```

Met Gln Glu Gln Met Asp Glu Ala
245

(2) INFORMATION FOR SEQ ID NO: 11:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 659 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA to mRNA
- (iii) HYPOTHETICAL: NO
- (iii) ANTI-SENSE: NO
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Arabidopsis thaliana
- (vii) IMMEDIATE SOURCE:
 - (B) CLONE: 4A24

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

CGCCGCCACC GCGATGTACG TGATCTACCA CCCTCGTCCG CCGTCGTTCT CCGTCCCGTC	60
AATAAGAATC AGCCGCGTGA ACCTAACAAC CTCCTCTGAT TCCTCCGTCT CTCATCTCTC	120
TTCTTCTTTC AACTTCACTC TAATCTCAGA GAATCCAAAC CAACACCTCT CTTTCTCTTA	180
CGATCCTTTC ACCGTCACCG TTAATTGAGC TAAATCCGGT ACGATGCTCG GTAACGGAAC	240
TGTTCCCTGCT TTCTTCAGCG ATAACGGTAA CAAAACTTCG TTTCACGGCG TGATCGCTAC	300
GTCTACAGCG GCGCGTGAGT TAGATCCGGA TGAAGCTAAG CATCTGAGAT CAGATCTGAC	360
GCGCGCGCGT GTAGGATATG AGATCGAGAT GAGAACTAAA GTGAAGATGA TAATGGGGAA	420
GCTGAAGAGT GAAGGAGTAG AGATCAAAGT GACATGTTGA AGGATTGAA GGAACATATAC	480
CAAAAGGTAA AACTCCAATT GTAGCTACTT CTAAAAAAC TAAGTGTAAG TCTGATCTTA	540
GTGTCAAGTC TGGAAATGGA TTTCTAAAGG AATTTGATAA TTTCACATTG AAATTCTATA	600
TATCTCTCTT TTTCTCTGGA TTTGTGAAAC TTTGGATGAT CAAAGAATTC TTCATTGTC	659

(2) INFORMATION FOR SEQ ID NO: 12:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 174 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single

- 36 -

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Arabidopsis thaliana

(vii) IMMEDIATE SOURCE:

(B) CLONE: 4A24

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

Arg	Ile	Cys	Cys	Cys	Cys	Phe	Trp	Ser	Ile	Leu	Ile	Ile	Leu	Ile	Leu
1				5					10					15	

Ala	Leu	Met	Thr	Ala	Ile	Ala	Ala	Thr	Ala	Met	Tyr	Val	Ile	Tyr	His
			20					25					30		

Pro	Arg	Pro	Pro	Ser	Phe	Ser	Val	Pro	Ser	Ile	Arg	Ile	Ser	Arg	Val
		35					40					45			

Asn	Leu	Thr	Thr	Ser	Ser	Asp	Ser	Ser	Val	Ser	His	Leu	Ser	Ser	Phe
50						55					60				

Phe	Asn	Phe	Thr	Leu	Ile	Ser	Glu	Asn	Pro	Asn	Gln	His	Leu	Ser	Phe
65				70					75					80	

Ser	Tyr	Asp	Pro	Phe	Thr	Val	Thr	Val	Asn	Ser	Ala	Lys	Ser	Gly	Thr
				85					90					95	

Met	Leu	Gly	Asn	Gly	Thr	Val	Pro	Ala	Phe	Phe	Ser	Asp	Asn	Gly	Asn
			100					105					110		

Lys	Thr	Ser	Phe	His	Gly	Val	Ile	Ala	Thr	Ser	Thr	Ala	Ala	Arg	Glu
		115					120					125			

Leu	Asp	Pro	Asp	Glu	Ala	Lys	His	Leu	Arg	Ser	Asp	Leu	Thr	Arg	Ala
	130					135					140				

Arg	Val	Gly	Tyr	Glu	Ile	Glu	Met	Arg	Thr	Lys	Val	Lys	Met	Ile	Met
145					150					155				160	

Gly	Lys	Leu	Lys	Ser	Glu	Gly	Val	Glu	Ile	Lys	Val	Thr	Cys
			165					170					

(2) INFORMATION FOR SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 584 base pairs

(B) TYPE: nucleic acid

- 37 -

(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: cDNA to mRNA
(iii) HYPOTHETICAL: NO
(iii) ANTI-SENSE: NO
(vi) ORIGINAL SOURCE:
 (A) ORGANISM: Arabidopsis thaliana
(vii) IMMEDIATE SOURCE:
 (B) CLONE: 3B76

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

CCTCCA	ACTC	CAGGCC	CAGCC	AACAAA	AAGAA	CCTAC	ATTTA	TTCCAG	TGGT	TGTT	TGGT	CCTT	60
TTGG	ACTC	AA	GTGGG	AA	GA	CATT	ACTCTT	TCCT	CTGT	TTC	ATT	TGAT	120
TGG	ACTC	AA	GTGGG	AA	GA	CATT	ACTCTT	TCCT	CTGT	TTC	ATT	TGAT	180
ACC	ATTT	CAG	GCAG	CAG	CAC	A	TA	CTT	CGA	GTG	ACA	AG	240
AT	CT	TAT	ACC	AGAA	AG	ACCT	GTT	CCG	TCCC	TATTT	AG	GGG	300
AT	CT	TAT	ACC	AGAA	AG	ACCT	GTT	CCG	TCCC	TATTT	AG	GGG	360
GAA	ACT	GAT	C	TCT	TA	TGA	TG	ACT	TAT	TTC	TT	CCT	420
TA	AG	GT	GG	G	AGG	CC	GT	TCA	AGT	TCT	GG	CA	480
TA	AG	GT	GG	G	AGG	CC	GT	TCA	AGT	TCT	GG	CA	540
TT	CT	CT	G	AT	AG	CA	AA	AG	CA	TA	AA	CA	600
TT	CT	CT	G	AT	AG	CA	AA	AG	CA	TA	AA	CA	660
GG	AG	AT	AA	TGG	AC	AT	GAT	GG	CC	GT	GG	CG	720
GG	AG	AT	AA	TGG	AC	AT	GAT	GG	CC	GT	GG	CG	780
TT	TG	TAC	GAA	AG	CAG	CT	TGC	AT	CT	GAA	CT	T	840
TT	TG	TAC	GAA	AG	CAG	CT	TGC	AT	CT	GAA	CT	T	900

(2) INFORMATION FOR SEQ ID NO: 14:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 283 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
(ii) MOLECULE TYPE: protein
(iii) HYPOTHETICAL: NO
(iii) ANTI-SENSE: NO
(vi) ORIGINAL SOURCE:
 (A) ORGANISM: Arabidopsis thaliana

- 38 -

(vii) IMMEDIATE SOURCE:
(B) CLONE: 3B76

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

Pro	Pro	Thr	Pro	Gly	Gln	Pro	Thr	Lys	Glu	Pro	Thr	Phe	Ile	Pro	Val	1	5	10	15
Val	Val	Gly	Leu	Leu	Asp	Ser	Ser	Gly	Lys	Asp	Ile	Thr	Leu	Ser	Ser	20	25	30	
Val	His	Tyr	Asp	Gly	Thr	Val	Gln	Thr	Ile	Thr	Gly	Ser	Ser	Thr	Ile	35	40	45	
Leu	Arg	Val	Thr	Lys	Lys	Gln	Glu	Glu	Phe	Val	Phe	Ser	Asp	Ile	Pro	50	55	60	
Glu	Arg	Pro	Val	Pro	Ser	Leu	Phe	Arg	Gly	Phe	Ser	Ala	Pro	Val	Arg	65	70	75	80
Val	Glu	Thr	Asp	Leu	Ser	Asn	Asp	Asp	Leu	Phe	Phe	Leu	Leu	Ala	His	85	90	95	
Asp	Ser	Asp	Glu	Phe	Asn	Arg	Trp	Glu	Ala	Gly	Gln	Val	Leu	Ala	Arg	100	105	110	
Lys	Leu	Met	Leu	Asn	Leu	Val	Ser	Asp	Phe	Gln	Gln	Asn	Lys	Pro	Leu	115	120	125	
Ala	Leu	Asn	Pro	Lys	Phe	Val	Gln	Gly	Leu	Gly	Ser	Val	Leu	Ser	Asp	130	135	140	
Ser	Ser	Leu	Asp	Lys	Glu	Phe	Ile	Ala	Lys	Ala	Ile	Thr	Leu	Pro	Gly	145	150	155	160
Glu	Gly	Glu	Ile	Met	Asp	Met	Met	Ala	Val	Ala	Asp	Pro	Asp	Ala	Val	165	170	175	
His	Ala	Val	Arg	Lys	Phe	Val	Arg	Lys	Gln	Leu	Ala	Ser	Glu	Leu	Lys	180	185	190	
Glu	Glu	Leu	Lys	Ile	Val	Glu	Asn	Asn	Arg	Ser	Thr	Glu	Ala	Tyr	Val	195	200	205	
Phe	Asp	His	Ser	Asn	Met	Ala	Arg	Arg	Ala	Leu	Lys	Asn	Thr	Ala	Leu	210	215	220	
Ala	Tyr	Leu	Ala	Ser	Leu	Glu	Asp	Pro	Ala	Tyr	Met	Gly	Thr	Cys	Thr	225	230	235	240
Glu	Arg	Ile	Gln	Gly	Gly	His	Gln	Phe	Asp	Arg	Pro	Ile	Cys	Cys	Phe	245	250	255	
Gly	Thr	Leu	Ser	Gln	Asn	Pro	Gly	Lys	Thr	Arg	Glu	Arg	Thr	Phe	Leu				

- 39 -

260	265	270
Pro Asp Phe Tyr Glu Gln Val Ala Gly Thr Ile		
275	280	

(2) INFORMATION FOR SEQ ID NO: 15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 534 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: *Arabidopsis thaliana*

(vii) IMMEDIATE SOURCE:

(B) CLONE: 4A5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

ACCAGGAGGG GAAAAAGTCT TACCCCATGG ACATCCCGGG GATTGAGTGT TACCCGAAAA	60
GGATGAAGAA TGGTATTCCT CCGTCGTGGA CCCCATGCAC CCATTGGGAA AGCCGTGTGG	120
CGTTTTCTTT CAGGGATGAT AGAAAAGTGC TCCCTTGGA TGGAAAGGAG GAGCCTTTAC	180
TGGTAGTGGC CGATAGGGTG AGGAATGTTG TGGAGGCTGA TGACGGGTAT TATCTCGTGG	240
TGGCTGAGAA CGGACTTAAG CTAGAGAAAG GATCAGATTT GAAGGCGAGA GAGGTGAAGG	300
AGAGTTTAGG GATGGTTGTT TTGGTGGTGA GGCCGCCAAG AGAAGATGAT GATGATTGGC	360
AGACAAGTCA TCAGAACTGG GACTGAATTA ATAGAATCAA TACTCATATG CTGTAACTGA	420
TTACGGAGTC ATCATGGTCA TGTAATAATTT TTGGATAAAG GTGGTAACTT TTTGTTCTAA	480
GATACAATCA GAAACAGAGC AATATTTTTTC TCTAAAAAAA AAAAAAAAAA AAAA	534

(2) INFORMATION FOR SEQ ID NO: 16:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 119 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

- 40 -

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: *Arabidopsis thaliana*

(vii) IMMEDIATE SOURCE:

(B) CLONE: 4A5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

Met Asp Ile Pro Gly Ile Glu Cys Tyr Pro Lys Arg Met Lys Asn Gly
 1 5 10 15

Ile Pro Pro Ser Trp Thr Pro Cys Thr His Trp Glu Ser Arg Val Ala
 20 25 30

Phe Ser Phe Arg Asp Asp Arg Lys Val Leu Pro Trp Asp Gly Lys Glu
 35 40 45

Glu Pro Leu Leu Val Val Ala Asp Arg Val Arg Asn Val Val Glu Ala
 50 55 60

Asp Asp Gly Tyr Tyr Leu Val Val Ala Glu Asn Gly Leu Lys Leu Glu
 65 70 75 80

Lys Gly Ser Asp Leu Lys Ala Arg Glu Val Lys Glu Ser Leu Gly Met
 85 90 95

Val Val Leu Val Val Arg Pro Pro Arg Glu Asp Asp Asp Asp Trp Gln
 100 105 110

Thr Ser His Gln Asn Trp Asp
 115

(2) INFORMATION FOR SEQ ID NO: 17:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vii) IMMEDIATE SOURCE:

(B) CLONE: primer V6

- 41 -

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

ATGCTTTGCA TAACTTTGAG G

21

(2) INFORMATION FOR SEQ ID NO: 18:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 17 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vii) IMMEDIATE SOURCE:

(B) CLONE: primer T7

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

AATACGACTC ACTATAG

17

What we claim is:

1. A method for increasing the probability of vegetative reproduction of a new plant generation comprising transgenically expressing a gene encoding a protein acting in the signal transduction cascade triggered by the Somatic Embryogenesis Receptor Kinase (SERK).
2. A method according to claim 1, wherein the encoded protein physically interacts with SERK.
3. The method according to claim 2, wherein the protein is a member of the family of Squamosa-promoter Binding Protein (SBP) transcription factors or 14-3-3 type lambda proteins.
4. The method according to claim 2, wherein the protein has the amino acid sequence given in SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, SEQ ID NO: 10, SEQ ID NO: 12, SEQ ID NO: 14, or SEQ ID NO: 16, or an amino acid sequence having a component sequence of at least 150 amino acids length which after alignment reveals at least 40% identity with SEQ ID NO: 12 or SEQ ID NO: 16.
5. The method according to claim 1 increasing the probability of vegetative reproduction through seeds (apomixis).
6. The method according to claim 5, wherein the seeds result from non-gametophytic apomixis.
7. The method according to claim 5, wherein the encoded protein is transgenically expressed in the vicinity of the embryo sac.
8. The method according to claim 1 increasing the probability of *in vitro* somatic embryogenesis.
9. The method according to claim 1, wherein expression of the gene is under control of the SERK gene promoter, the carrot chitinase DcEP3-1 gene promoter, the *Arabidopsis*

AtChitIV gene promoter, The *Arabidopsis* LTP-1 gene promoter, The *Arabidopsis* bel-1 gene promoter, the petunia fbp-7 gene promoter, the *Arabidopsis* ANT gene promoter or the promoter of the O126 gene of *Phalaenopsis*.

10. A gene encoding a protein having the amino acid sequence given in SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, SEQ ID NO: 10, SEQ ID NO: 12, SEQ ID NO: 14, or SEQ ID NO: 16, or an amino acid sequence having a component sequence of at least 150 amino acids length which after alignment reveals at least 40% sequence identity with SEQ ID NO: 12 or SEQ ID NO: 16.
11. A gene according to claim 10 having the nucleotide sequence given in SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 9, SEQ ID NO: 11, SEQ ID NO: 13, or SEQ ID NO: 15.
12. A gene according to claim 10 wherein the nucleotide sequence is modified in that known mRNA instability motifs or polyadenylation signals are removed and/or codons which are preferred by the plant into which the DNA is to be inserted are used.
13. A plant or plant cell transgenically expressing the gene according to any one of claims 10-12.
14. A plant or plant cell obtainable by the method of claim 1.

1. The first part of the document is a list of the names of the persons who were present at the meeting. The names are listed in alphabetical order.

2. The second part of the document is a list of the topics that were discussed at the meeting. The topics are listed in alphabetical order.